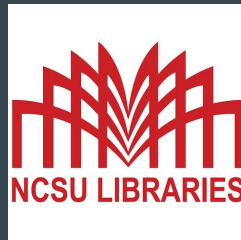


# Data Cleaning with Open Refine



Alison Blaine, [ablaine@ncsu.edu](mailto:ablaine@ncsu.edu)  
Jennifer Garrett, [jigarret@ncsu.edu](mailto:jigarret@ncsu.edu)



# Data Visualization Services

[lib.ncsu.edu/do/visualization](http://lib.ncsu.edu/do/visualization)



FIND

GET HELP

SERVICES

LIBRARIES

ABOUT

Search books, articles, journals, website



ASK US | MY ACCOUNT | HOURS | FAQ | LOG OUT | CHAT NOW

[← Back to Research Support](#)

## NEED HELP?

› Ask a question or  
request a consultation



Alison Blaine



Karen Ciccone



## Visualization

A good data visualization can expose new patterns and relationships in your data and help you communicate about it more clearly. We can help you:

- › use visualization spaces and technologies
- › use GIS and online mapping software
- › learn how to create more effective data visualizations
- › visualize research impact

## Visualization Spaces

The Libraries provides spaces and technologies for high-resolution, immersive, 3D, and interactive displays of research data. We provide [boilerplate facilities descriptions](#) for inclusion in grant proposals. [Contact us](#) for more information and assistance.

## WORKSHOPS

JAN

15

Getting Started with Data  
Visualization: Tools for  
Research!

11:00 AM to 12:00 PM

JAN

15

Creative Coding Group

12:15 PM to 1:15 PM

JAN

27

Visualize Your Data with  
Tableau

3:00 PM to 4:00 PM

JAN

29

Data Cleaning and  
Analysis Tips and Tools:  
Excel and Open Refine

11:00 AM to 12:00 PM

[View all workshops ›](#)

# Data Visualization Process

- getting data
- getting to know the data
- setting goals
- determining if a visualization is needed
- deciding what to visualize
- cleaning
- visualizing
- potentially re-visualizing it with a different tool
- exporting/embedding

# Data cleaning

Cleaning data means removing problematic characters, whitespaces, symbols, errors, missing elements and inconsistencies.

C
Raleigh
Raleigh, NC
Raleigh, N.C.
Raleigh, NC 27695
Raleigh, NC USA

Tools:

- Open Refine
- ASAP Utilities for Excel
- Data cleaning using code in R or SAS

# Open Refine

- Formerly owned by Google, now maintained by independent developers.
- Free, open source tool that makes it easy to clean and transform large data sets.
- You can find problems in your data set without knowing about them in advance.
- Requires downloading the software in order to use (<http://openrefine.org/download.html>)
- GREL documentation:  
<https://github.com/OpenRefine/OpenRefine/wiki/GREL-Functions>

Today's workshop materials

[go.ncsu.edu/refine](https://go.ncsu.edu/refine)